
Procedural Considerations Concerning Data Collectors and Types of Data Obtained in Household Interview Surveys

LAURENCE G. BRANCH, PhD

SEVERAL VARIABLES CAN AFFECT the reliability of data acquired in interview surveys. These variables include the background of the data collector and the form or types of data collected. The household survey is an increasingly popular method for obtaining baseline assessments of the health status and health care needs of the people of a region or health service area. These assessments then serve as building blocks in the structure of planning for health care, or they can be used to evaluate the effects of programs already in progress in a target population. This examination of various options in data collection focuses on the choices that affect the reliability of the resulting data.

Since it is axiomatic that reliability must precede validity, the issue of which data are correct (or valid) is irrelevant if the data are not first reliable. This examination of various options in data collection is not concerned with the validity issue but rather it is focused on the question of which method produces more reliable information.

Feldman has summarized much of the research on the general factors influencing reliability in the process of demarcating the strengths and weaknesses of the

household survey for the collection of health event information (1). This summary provides an excellent overview for understanding the general problem of reliability. Two specific methodological studies, one by Elinson and Trussell (2) and the other by Madow (3), highlight the basic issue of how specific procedures can influence both the reliability and validity of the findings.

Focusing specifically on two conditions, arthritis and heart disease, Elinson and Trussell examined the degree of correspondence between diagnostic information self-reported during personal interviews and that obtained by clinical examinations. The two procedures did not produce the kind of consistency necessary to conclude that both can produce reliable information. About one in four of the self-reported conditions was not substantiated by the clinical examination, while about two out of three conditions found in the more objective clinical examinations had not been reported in the more subjective survey assessment. It is possible, however, that some of the respondents did not know of the conditions found in the examination.

Madow compared the completeness and accuracy with which chronic conditions are reported by health plan enrollees in interviews to the information obtained from physicians' records. Again the correspondence was disappointing: about 40 percent of the self-reported information was missing from the physicians' records, while almost half of the conditions listed in the physicians' records were not mentioned by the patient in the interview.

In this report two types of data collection (self-reporting and clinical judgments) were studied as they related to two types of data collectors (trained lay interviewers and nurse interviewers). Thus, nurses made clinical judgments and recorded self-reported informa-

Dr. Branch is an assistant professor in the Department of Preventive and Social Medicine, Harvard Medical School. The survey described in the paper was supported in part by a contract with the Massachusetts Department of Public Health and by grants 90-A-641/01 and 90-A-641/02 from the Administration on Aging, Department of Health, Education, and Welfare, when Dr. Branch was a program director at the Center for Survey Research. The center is a facility of the University of Massachusetts and the Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University.

Tearsheet requests to Dr. Laurence G. Branch, Department of Preventive and Social Medicine, Harvard Medical School, 641 Huntington Ave., Boston, Mass., 02115.

tion; lay interviewers made clinical judgments and recorded self-reported information.

The degree to which these different options were or were not interchangeable in the production of consistent data became the test for reliability. It would obviously seem more desirable to use lay interviewers rather than nurses in household health surveys since the lay interviewers are more readily available and less expensive, but the critical issue is whether the information they would provide would be as reliable as that provided by nurses. Can lay interviewers, accustomed to working with structured questionnaires, make sensitive clinical judgments consistent with those made by trained nurses who have clinical skills to rely on? Does the professional expertise of the nurse lend any special talent to the structured interview? Can the nurse elicit better self-reporting by careful probes based on her or his special knowledge?

Methods

Questionnaire. The structured questionnaire contained 22 factual demographic items, 27 attitudinal items, 44 health event or health condition items, 17 items based on a modification of the Katz Index of Activities of Daily Living (4), and 25 judgments by the interviewer of the respondent's health status and functional abilities. Fourteen of these judgments replicated those used to generate data from the household sample in a Monroe County, N.Y., study (5) that assessed the health needs of the aged. Seven other judgments replicated those made by the Medical Review Team of Massachusetts in its routine reviews of institutionalized persons for the appropriateness of their placement and treatment. The other four were devised for this particular investigation.

Respondents. The respondents, 122 elderly persons, included 57 residents of 8 nursing homes and 65 residents of public housing in neighborhoods close to the nursing homes.

Interviewers. Eleven experienced female interviewers from an academic survey research group and seven female nurses were chosen to administer the questionnaire. The nurses, all of whom had field experience as visiting nurses in Massachusetts, participated in a 4-day training session; during it the survey research staff instructed them in established social science interviewing methods for collecting self-reported information using a structured questionnaire. The 11 lay interviewers, already skilled in such techniques, joined the nurses in the last 2 days of the training session for briefing on the specific questionnaire and on the background of the study.

Field procedures. A nurse and a lay interviewer, paired as a team, were present at every interview. Using a counterbalanced procedure, each recorded the respondents' answers separately while alternating roles as active interviewers. Thus, the lay interviewer conducted half of the interviews and recorded the responses while the nurse observed passively and recorded her version of the responses. The nurse then conducted the interviews and recorded. After the standardized questionnaire had been completed, the passive partner in the team had an opportunity to probe any areas in the questionnaire which she judged needed further exploration. This procedure was particularly designed to enable the visiting nurses to take advantage of their professional health experience by identifying discrepancies between what respondents claimed their condition to be

and what the nurse believed it to be. After this followup probing was completed, the nurse and the professional interviewer independently filled out ratings of the respondent's disability level.

Alternating the active recorder did not have any systematic influence on either the type of difference in the material recorded by the two types of collectors (that is, one did not consistently record more problems or disabilities than the other) or on the form of such differences (that is, one did not more frequently leave an item blank or record an answer differently from that recorded by her counterpart). Consequently, this control variable is not used as an analytic variable in the analyses.

In addition, the floor nurses responsible for the 57 nursing home residents were asked to fill out a shortened version of the questionnaire focusing on the health care needs of the person and replicating the ratings made by the interviewers. The floor nurses were given no special training for this task. Floor nurses completed and returned 54 questionnaires.

The survey was conducted in September and October 1974. Each interview with a respondent lasted approximately 50 minutes.

Results and Discussion

Statistically testing whether an obtained difference rate is significantly different from zero would assume that total interchangeability between the types of data collectors or between the types of information collected

is possible. Such an assumption is unrealistic. In the context of this study, it is more realistic to assume that the demographic items are the most straightforward and, therefore, are subject to the least amount of potential discrepancy between methods. Accordingly, the following analyses were undertaken to determine if any other items or procedures produced significantly more discrepancy than the difference rate between interviewers and visiting nurses for the factual items.

In such analyses, two kinds of difference rates are commonly investigated (6): the gross difference rate and the net difference rate. The gross difference rate is the sum of all the disagreements in the classification of respondents divided by the total number of respondents classified. The gross difference rate includes differences in classification in both directions independent of whether the differences may cancel each other out. The net difference rate is the sum of the differences which do not cancel each other out divided by the total number of respondents classified. The net difference rate is therefore an indicator of the error in the data which would confound projections at the aggregate level.

Of course, from the point of view of the individual, any misclassification has practical implications, but misclassifying five persons as needing help when they do not and misclassifying five others in the opposite direction (not needing help when they do) does not alter the status of the group as a whole and, therefore, has no practical significance in planning for the aggregate.

Table 1. Differences among information collectors for self-reported and judgmental information

Type of Information	Gross differences			Net differences		
	Percent disagreeing	Significantly different from rate for factual items		Percent disagreeing	Significantly different from rate for factual items	
		Level	z^1		Level	z^1
Self-reported:						
Factual items	6	5
Attitudinal items	8	.001	4.79	6	.01	2.92
Health event and health condition items	5	NS	1.87	3	.001	4.91
Activities of daily living items (modified Katz scale) ..	4	.001	4.59	3	.001	4.54
Judgmental:						
Items from the Monroe County, N.Y., study	13	.001	12.39	8	.001	6.81
Judgments to assess respondent's orientation to physical reality (hand-face test)	13	.001	10.70	5	NS	0.82
Items from the Massachusetts Long-Term Care Patient Assessment Form	15	.001	13.04	7	.001	4.12

¹ The test of significance between 2 proportions was used. No correction for related samples was taken for 2 reasons. First, the logic of comparing the difference found in some types of items to the difference found with factual items parallels the logic which suggests subtracting a correction term in the denominator for the test with related samples; namely, that some amount of within-subject error is going to exist and ought to be canceled in the actual statistical test. Second,

the practical effect of both options (incorporating a correction term with related samples or comparing the difference of 1 type of item to the difference found with factual items) is to make a more conservative test which requires a larger difference in order to conclude a statistically significant difference.

NOTE: NS = not significant.

Comparing the reliability of self-reported information with clinical judgments. Table 1 indicates the rate of differences between lay interviewers and visiting nurses in recording the four types of self-reported information and the three types of judgments. There was a 6 percent gross difference rate for factual items. This percentage was the arbitrary standard of minimum, tolerable differences. The gross difference rates for all four types of self-reported information are fairly homogeneous, ranging from a low of 4 percent on the 17-item activities of daily living (ADL) index to a high of 8 percent for attitudinal items. The consistency of the four types of self-reported information as recorded by each team member is comparable and quite high.

In contrast, the gross difference rates between the team members on the three types of judgments homogeneously clustered between 13 and 15 percent, or about twice as great as the differences obtained for self-reported information, indicating that judgments are considerably less reliable than self-reported information. This finding has disquieting implications for the validity of generalizations based on studies which derive their data solely from judgmental observations of interviewers or nurses. (Although one might counter that these judgments only appear to be comparably unreliable because the inadequate assessments of the lay interviewers dilute the more expert opinions of the visiting nurses, this objection is not supported subsequently in a comparison of the judgments made by visiting nurses and floor nurses.)

The net difference rates likewise suggest that judgmental information is, in general, less reliable than self-reported information. Therefore, even when one considers the practical implications of planning for the community, self-reported information is more consistent than judgmental data.

Assessing the utility of a health background for the data collector. The explanation of the larger gross and net difference rates for judgments than for self-reported information cannot necessarily, at this point, be attributed solely to the items themselves. Certainly the differences in the backgrounds of the lay interviewers and nurses could affect the gross and net difference rates. As a partial test of the influence of health training background in making the clinical judgments, gross difference rates among lay interviewers, visiting nurses, and floor nurses were compared on 14 independent judgments of the 54 residents of nursing homes. Seven of the 14 replicate the Massachusetts Long Term Care Patient Assessment Form; the other 7 are from the Monroe County study. They include ambulation, con-

tinence, and mental status. (Judgments for specific institutional care, which were moot for this institutionalized group of respondents, were excluded.)

The data in table 2 indicate that all three sources agreed in slightly more than half of the judgments, a percentage that leaves a considerable amount of disagreement. The rate that the floor nurses differed from visiting nurses or from lay interviewers was twice as great as the rate at which visiting nurses and lay interviewers disagreed. Apparently health background per se did not mitigate the rate of difference between visiting nurses and lay interviewers because the two types of nurses disagreed twice as frequently on a set of judgments as the visiting nurses and lay interviewers did on the same judgments.

This analysis can only claim to be a "partial test" of reliability because the two groups of nurses differed on another possibly significant but uncontrolled variable, namely their length of contact with the respondents. Visiting nurses spent approximately 1 hour with the respondents, but the floor nurses had a considerably longer period of contact. The floor nurses also had access to the patients' medical charts and a general, long-term knowledge of the clinical diagnoses which the visiting nurses lacked. The effects of these uncontrolled variables cannot be quantified for this analysis.

Table 2. Comparing differences for two types of judgments recorded by lay interviewers, visiting nurses, and floor nurses, in percentages

Groups compared	Judgment Items from—	
	Monroe County, N.Y., study	Massachusetts Long-Term Care Patient Assessment Form
Gross difference rates among all 3:		
All 3 agree	56	52
Lay interviewer differs from other 2	8	4
Visiting nurse differs from other 2	4	5
Floor nurse differs from other 2	24	30
All 3 disagree	8	9
Cumulative differences between visiting nurses and floor nurses	36	44
Cumulative differences between floor nurses and lay interviewers ..	40	43
Cumulative differences between visiting nurses and lay interviewers ..	20	18

The data in table 2 suggest that the judgmental items themselves generated unreliability, rather than the health backgrounds of the judges. This finding is further supported by the data in table 3 which indicate that visiting nurses and lay interviewers are interchangeable for collecting self-reported health events information. Apparently the reliability of such information rests with the type and form of information being gathered and not with the background of the interviewer.

Self-reported information and clinical judgments on activities of daily living. Since health status data are frequently used to categorize people into convenient aggregates that help to determine health policy, it is important to be aware that the method used to collect these data can lead to different categorizations of the same person. Self-reported information from this specific group of elderly respondents concerning their independence in ambulation, bathing, dressing, and feeding showed a greater degree of personal independence than was indicated by interviewers' clinical judgments of them.

The data suggesting this generalization come from several sources. Consider first the gross and net difference rates for each source (lay interviewers, visiting nurses, floor nurses) using two different procedures with the same respondents (table 4). For the lay interviewers and the visiting nurses, the two procedures were self-reported information compared to their judgments based on the Massachusetts Long Term Care Patient Assessment Form; for the floor nurses, the two procedures were both clinical judgments with different formats—the self-report items revised into judgment items and the items in the long term care assessment form. Bear in mind that a "difference" reported in table 4

Table 3. Comparing differences for health event and health condition items recorded by lay interviewers, visiting nurses, and floor nurses

Groups compared	Percent
Gross difference rates among all 3:	
All 3 agree	78
Lay interviewer differs from other 2	3
Visiting nurse differs from other 2	1
Floor nurse differs from other 2	17
All 3 disagree	1
Cumulative differences between visiting nurses and floor nurses	19
Cumulative differences between floor nurses and lay interviewers	21
Cumulative differences between visiting nurses and lay interviewers	5

means that the person was classed as independent with respect to the specific function by one procedure and as requiring some kind of assistance by the other.

Focusing first on the differences by source in table 4, one notices that lay interviewers had difficulty aligning their judgment about ambulation with the respondent's perception; the respondent decidedly more often was of the opinion that he or she was independent. Visiting nurses also judged the respondents to be more dependent in ambulation than the respondents themselves reported. Visiting nurses also differed in their judgments with the respondents' reported level of independence in bathing.

Table 4. Differences between self-reported information and judgments about activities of daily living made by lay interviewers, visiting nurses, and floor nurses

Activity	Disagreement (In percentages) by—		
	Lay interviewer	Visiting nurse	Floor nurse
Ambulation:			
Gross difference rate	15	10	8
Net difference rate	14	7	3
Self-reported independence, judged dependent	14	8	5
Self-reported dependence, judged independent	1	2	3
Bathing:			
Gross difference rate	7	15	12
Net difference rate	4	2	7
Self-reported dependence, judged dependent	6	8	10
Self-reported dependence, judged independent	2	7	2
Dressing:			
Gross difference rate	7	2	2
Net difference rate	7	2	2
Self-reported independence, judged dependent	7	2	2
Self-reported dependence, judged independent	0	0	0
Feeding:			
Gross difference rate	7	2	0
Net difference rate	3	2	0
Self-reported independence, judged dependent	5	2	0
Self-reported dependence, judged independent	2	0	0
Cumulative:			
Gross difference rate	9	7	6
Net difference rate	7	3	3
Self-reported independence, judged dependent	8	5	4
Self-reported dependence, judged independent	1	2	1
Reported independent	88	87	79
Reported dependent	12	13	21
Judged independent	81	84	76
Judged dependent	19	16	24

The cumulative difference rates across all four activities indicated a consistent pattern when differences occurred—the self-reported information leaned significantly in the direction of personal independence ($z=7.53$ for lay interviewers and 3.60 for visiting nurses). These results indicate that the respondents reported that they are better able to take care of themselves than the interviewers or nurses thought. The pattern of differences found for the floor nurses were also interesting. They were inconsistent in their judgments of the patient's independence in bathing as a specific activity, and they judged significantly ($z=2.45$) more dependence across all activities with the Massachusetts Long Term Care Patient Assessment Form than with the modified self-report items. It is no surprise that the format and sequencing of clinical judgments influence judgmental data as much as format and sequencing influence self-reported data.

Four generalizations can be drawn from the foregoing analysis of the data:

1. Self-reported information produces significantly more consistent data across different types of data collectors than do clinical judgments.
2. Lay interviewers and visiting nurses can be equally well trained to produce consistent results with self-reported information.
3. The health training background of the information collector does not increase the consistency of the information.
4. Consequently, the recommended procedure for

maximizing the basic consistency of health planning information is a structured interview designed to obtain self-reported information, rather than clinical judgments. The decision to use either experienced lay interviewers or visiting nurses can be dictated solely by relative costs and availability. Self-reported information is likely to indicate more independence or better functioning in certain areas than clinical judgments, but in this study it was not possible to determine which data more accurately reflect the reality.

References

1. Feldman, J.: The household interview survey as a technique for the collection of morbidity data. *J Chronic Dis* 11: 535-557 (1960).
2. Elinson, J., and Trussell, R.: Some factors relating to degree of correspondence for diagnostic information as obtained by household interviews and clinical examinations. *Am J Public Health* 47: 311-321 (1957).
3. Madow, W.: Net differences in interview data on chronic conditions and information derived from medical records. DHEW Publication No. (HSM) 73-1331, Series 2, No. 57, Rockville, Md., 1973.
4. Katz, S., Downs, T., Cash, M., and Grotz, R.: Progress in development of the Index of ADL. *Gerontologist* 10: 20-30 (1970).
5. Berg, R. L., Browing, F. E., Hill, J. G., and Wenkert, W.: Assessing the health care needs of the aged. *Health Serv Res* 5: 36-59 (1970).
6. Bureau of the Census: The current population survey reinterview program, January 1961 through December 1966. Technical Paper No. 19, U.S. Government Printing Office, Washington, D.C., 1968.

SYNOPSIS

BRANCH, LAURENCE G. (Harvard Medical School): *Procedural considerations concerning data collectors and types of data obtained in household interview surveys. Public Health Reports, Vol. 94, March-April 1979, pp. 136-141.*

The administrative decision to gather baseline data on health care or social service needs as a foundation for health planning often generates discussions about appropriate procedures for data collection. Two significant issues frequently raised are (a) whether to use self-reported information from respondents or judgments about the health status and needs made by interviewers and (b) should health care professionals (usually nurses) be used to make the

judgments or even to collect the self-reported information regarding health events and conditions.

In a controlled experiment, teams of lay interviewers and visiting nurses interviewed 122 elderly respondents to obtain four types of self-reported information (factual demographic items, attitudinal items, health characteristics, and ability in activities of daily living). The nurse and lay interviewer, both present at each interview, recorded information separately, and they alternated roles as the active interviewer. They used a structured questionnaire and also made three types of clinical judgments after approximately 1 hour of personal contact with each respondent. In addition, floor nurses provided

two types of clinical judgments on 54 respondents who were residents of long-term care nursing homes.

Analysis of the extent of agreement among the three groups and with the residents' self-reports led to these conclusions. The recommended procedure for maximizing the reliability of data for health planning is to use structured interviews to get self-reported information rather than using clinical judgments made by the interviewers. An experienced lay interviewer can be trained to use a health survey instrument as well as a visiting nurse. Self-reported information is likely to indicate greater independence or better functioning in certain activities than would be obtained in clinical judgments.